In the Claims

1. (Currently Amended) A process for modeling numerical data for forecasting a phenomenon from a data set relating to the phenomenon, the process comprising:

collecting data ~~for development of a model~~ with a data acquisition module for describing the phenomenon;

processing the data ~~to enhance its exploitability~~ in a data preparation module to enhance exploitability of the data;

constructing [[a]] an initial model ~~by learning on the processed data~~ in a modeling module by learning on the processed data, the initial model having a fit to the data, robustness and parameters;

evaluating the fit and robustness of the ~~obtained~~ initial model in a performance analysis module; and

adjusting the model parameters in an optimization module to select ~~the~~ an optimal model ~~in an optimization module~~, wherein the optimal model is generated in the form of a D^{th} order polynomial ~~of the~~ derived from variables ~~used in~~ input ~~of~~ into the modeling module[[,]] by controlling ~~the~~ a trade-off between ~~the~~ learning accuracy and ~~the~~ learning stability ~~with the addition,~~ the trade-off being controlled by adding to ~~the~~ a covariance matrix ~~of~~ a perturbation ~~during calculation of the model~~ in the form of ~~the~~ a product of a scalar λ times a matrix H or in the form of a matrix H dependent on a vector of k parameters $\Lambda=(\lambda_1,\lambda_2, \ldots \lambda_k)$ during calculation of the optimal model, wherein ~~where~~ the order D of the polynomial and the scalar λ, or the vector of parameters Λ, are determined automatically during model adjustment by the optimization module by integrating an additional data partition step performed by a partition module, the data partition step comprising the step of ~~which consists in~~ constructing ~~two preferably disjoint subsets:~~ a first subset comprising training data used as a learning base for the modeling module and a second subset comprising generalization data ~~destined~~ used to build a model validity criterion to adjust the value of the order D, the scalar λ or the vector of parameters Λ ~~these parameters according to a model validity criterion obtained on data that did not participate in the training~~, and ~~where~~ wherein the matrix H is a positive defined matrix of dimensions equal to ~~the~~ a number p of ~~input~~ variables input into the modeling module, plus one.

2. (Original) The data modeling process according to Claim 1, wherein the matrix H verifies the following conditions: H(i,i) is close to 1 for i between 1 and p, H(p+1,p+1) is close to 0 and H(i,j) is close to 0 for i different from j.

3. (Original) The data modeling process according to Claim 1, wherein the matrix H verifies the following conditions: H(i,i) is close to a for i between 1 and p, H(p+1,p+1) is close to b, H(i,j) is close to c for i different from j and a = b + c.

4. (Original) The data modeling process according to Claim 3, wherein the matrix H verifies the following additional conditions: a is close to 1-1/p, b is close to 1, c is close to -1/p.

5. (Original) The data modeling process according to Claim 1, wherein the matrix H verifies the following condition: H(p+1,p+1) is different from at least one of the terms H(i,i) for i between 1 and p.

6. (Currently Amended) The data modeling process according to Claim 1, wherein ~~base~~ the data partition step is performed by an operator using an external software program.

7. (Original) The data modeling process according to Claim 1, wherein the data partition module performs a pseudorandom sampling to construct two subsets from the base data.

8. (Currently Amended) The data modeling process according to Claim 1, wherein the data partition module performs a pseudorandom sampling to construct two subsets from the ~~base~~ data, while keeping the statistical representativeness of the input vectors in the two subsets.

9. (Currently Amended) The data modeling process according to Claim 1, wherein the ~~base~~ data partition module performs a sequential sampling to construct two subsets from the ~~base~~ data.

10. (Currently Amended) The data modeling process according to Claim 1, wherein the ~~base~~ data partition module performs a first split of the data into two subsets, with the first subset

comprising the training and generalization data and the second subset comprising the test data.

11. (Currently Amended) The data modeling process according to Claim 1, wherein the ~~base~~ data partition module performs a sampling of the type selecting at least one sample according to a law programmed in advance for generation of the training, generalization and/or test subsets.

12. (Currently Amended) The data modeling process according to Claim 1, wherein the optimization module selects the pair of parameters (D, λ) or (D, Λ) that minimizes one ~~or the other~~ of the following ~~quantities~~:

- mean error on the subset of the generalization data;
- weighted mean error on the subset of the generalization data;
- mean quadratic error on the subset of the generalization data;
- weighted mean quadratic error on the subset of the generalization data.

13. (Original) The data modeling process according to Claim 1, wherein the data preparation module performs a statistical normalization of columns of data.

14. (Currently Amended) The data modeling process according to Claim 1, wherein the data preparation module fills in missing data by one ~~or the other~~ of the following ~~quantities~~:

- mean of the value on a column (real type data);
- mean of the value on a subset of a column (real type data);
- most frequent value (Boolean or <<nominal>> type data);
- most frequent value on a subset of a column (Boolean or <<nominal>> type data);
- selection of a fixed substitution value;
- estimation of the substitution value on the basis of a modeling as a function of other variables.

15. (Original) The data modeling process according to Claim 1, wherein the data preparation module performs detection of outlying data according to one or more of the following criteria:

- data outside a range defined by an operator;

- data outside a range calculated by the system;

- Boolean or enumerated data whose number of occurrences is below a given threshold.

16. (Currently Amended) The data modeling process according to Claim 1, wherein the data preparation module performs a substitution of outliers by one ~~or the other~~ of the following ~~quantities~~:

- mean of the value on the column (real type data);
- mean of the value on a subset of the column (real type data);
- most frequent value (Boolean or <<nominal>> type data);
- most frequent value on a subset of the column (Boolean or <<nominal>> type data);
- selection of a fixed substitution value;
- estimation of the substitution value on the basis of a modeling as a function of other variables.

17. (Original) The data modeling process according to Claim 1, wherein the data preparation module performs a monovariable or multivariable polynomial development on all or part of the input data.

18. (Original) The data modeling process according to Claim 1, wherein the data preparation module performs a periodic development of the input data.

19. (Currently Amended) The data modeling process according to Claim 1, wherein the data preparation module performs an explicative development of ~~the~~ input data comprising dates ~~of date type~~.

20. (Original) The data modeling process according to Claim 1, wherein the data preparation module performs a change of coordinates, stemming from a principal components analysis with possible simplification.

21. (Original) The data modeling process according to Claim 1, wherein the data

preparation module performs one or more temporal shifts before or after all or part of a column containing time variables.

22. (Currently Amended) The data modeling process according to Claim 1, further comprising ~~exploration of~~ the step of exploring possible preparations by a preparation exploration module which uses a description of ~~the~~ possible preparations provided by the user and an exploration strategy based either on a pure performance criterion in training or in generalization, or on a ~~trade-off between these performances~~ combination of the pure performance criterion and ~~the~~ a capacity of the learning process obtained.

23. (Original) The data modeling process according to Claim 1, wherein the modeling further comprises model exploitation performed by an exploitation module which provides monovariable or multivariable polynomial formulas descriptive of the phenomenon.

24. (Original) The data modeling process according to Claim 18, wherein the modeling further comprises model exploitation performed by an exploitation module which provides periodic formulas descriptive of the phenomenon.

25. (Currently Amended) The data modeling process according to Claim 19, wherein the modeling further comprises model exploitation performed by an exploitation module which provides descriptive formulas of the phenomenon ~~containing~~, the descriptive formulas comprising date developments ~~in calendar indicators~~.

26. (Original) The data modeling process according to Claim 18, wherein the periodic development is a trigonometric development.

27. (Original) The data modeling process according to Claim 24, wherein periodic formulas descriptive of the phenomenon use trigonometric functions.

28. (Original) The data modeling process according to Claim 1, wherein the data preparation module performs one or more of the following actions on<<nominal>> data to reduce

the number of distinct values:

- calculation of the amount of information brought by each value;
- grouping with each other the values homogeneous in relation to the phenomenon under study;
- creation of a specific value regrouping all of elementary values not providing significant information on the phenomenon.

29. (Original) The data modeling process according to Claim 1, wherein the data preparation module regroups missing, outlying or exceptional data into one or more groups to apply a specific processing to them.

30. (Original) The data modeling process according to Claim 1, wherein the data preparation module performs encoding of "nominal" type data as tables of Boolean or real variables.

31. (Original) The data modeling process according to Claim 1, wherein the data preparation module calculates for each input variable its explicative power in relation to the phenomenon under study.

32. (Original) The data modeling process according to Claim 1, wherein the data preparation module uses segmentation algorithms.

33. (Original) The data modeling process according to Claim 1, wherein the data preparation module associates with each value of a <<nominal>> type datum a numerical value representative of the phenomenon under study.

34. (Original) The data modeling process according to Claim 1, wherein the data preparation module uses logical rules stemming from knowledge of the phenomena under study to encode dated data into Boolean values.

35. (Currently Amended) The data modeling process according to Claim 1, wherein the data preparation module ~~process flows by identifying~~ identifies periodic due dates and ~~applying to~~

~~them~~ applies management rules ~~appropriate~~ to each due date.

36. (Currently Amended) The data modeling process according to Claim 1, wherein the training[[,]] data subset and the generalization data subset ~~and forecasting spaces~~ are not disjointed.

37. (Currently Amended) The data modeling process according to Claim 1, further comprising storing and managing the data set and the formulas descriptive of the phenomenon in ~~wherein there is defined~~ a relational structure which contains variables, phenomena and models ~~for storing and managing the base data set and the formulas descriptive of the phenomenon~~.

38. (Currently Amended) A device for modeling numerical data for forecasting a phenomenon from a data sample relating to the phenomenon, the device comprising:

means for collecting input data;

means for processing the input data;

means for constructing [[a]] an initial model by learning on the processed data;

means for analyzing performances of the ~~obtained~~ initial model; and

means for optimizing the ~~obtained~~ initial model to generate an optimal model, wherein the optimal model is generated in the form of a D^{th} order polynomial ~~of the~~ derived from variables ~~used in~~ input ~~of~~ into the modeling module[[,]] by controlling ~~the~~ a trade-off between ~~the~~ learning accuracy and ~~the~~ learning stability ~~with the addition~~, the trade-off being controlled by adding to ~~the~~ a covariance matrix ~~of~~ a perturbation ~~during calculation of the model~~ in the form of ~~the~~ a product of a scalar λ times a matrix H or in the form of a matrix H dependent on a vector of k parameters $\Lambda=(\lambda_1,\lambda_2, \ldots \lambda_k)$ during calculation of the optimal model, wherein ~~where~~ the order D of the polynomial and the scalar λ, or the vector of parameters Λ, are determined automatically during model adjustment by the optimization module by integrating additional means for splitting the data so as to construct ~~two preferably disjoint subsets:~~ a first subset comprising training data used as a learning base for the modeling module and a second subset comprising generalization data ~~destined~~ used to build a model validity criterion to adjust the value of the order D, the scalar λ or the vector of

parameters Λ ~~these parameters according to a model validity criterion obtained on data that did not participate in the training~~, and ~~where~~ wherein the matrix H is a positive defined matrix of dimensions equal to ~~the~~ a number p of ~~input~~ variables input into the modeling module, plus one.

39. (New) The data modeling process of Claim 1, wherein the training data subset and the generalization data subset are disjointed.

40. (New) The device of Claim 38, wherein the training data subset and the generalization data subset are disjointed.